Parametric versus Nonparametric Goodness of Fit

Another View

H.Läuter and M.Nikulin

January 26, 1999

Department of Statistics, Institute of Mathematics, University of Potsdam, Germany
Mathematical Statistics, UFR MI2S, University Bordeaux 2, France
& Steklov Mathematical Institute, Saint Petersbourg, Russia

Abstract

We consider chi-squared type tests for testing the hypothesis H_0 that a density f of observations $X_1, ..., X_n$ lies in a parametric class of densities $\mathcal{F}$. We consider a version of chi-squared type test using kernel estimates for the density. The main result is, following Liero, Läuter and Konakov (1998), the derivation of the asymptotic behavior of the power of the test under Pitman and "sharp peak" type alternatives. The connection of the rate of convergence of these local alternatives, the bandwidth of the kernel estimator, the parametric estimator, the power of the test are studied.

Keywords and Phrases: Chi-squared test, goodness-of-fit test, density, kernel estimators, maximum likelihood estimator, local alternative, Pitman alternative, sharp peak alternative, asymptotic power.

AMS 1991 Subject Classification: Primary 62G10, 62G20.

The research for this paper was carried out within Sonderforschungsbereich 373 and was printed using funds made available by the Deutsche Forschungsgemeinschaft.

1. Introduction

Let $X_1, X_2, \cdots, X_n$ be a sample of i.i.d. random variables with density f. We wish to test whether f belongs to some parametric family $\mathcal{F}$ of density functions

$$\mathcal{F} = \{f : f(\cdot) = f(\cdot, \theta), \theta \in \Theta \subset \mathbf{R}^k\} \tag{1}$$

against the nonparametric alternative

$$f \notin \mathcal{F}.$$

We consider the kernel estimator $\hat{f}_n$ of the density function f,

$$\hat{f}_n(t) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h_n} K\left(\frac{t - X_i}{h_n}\right), \tag{2}$$

where the kernel function $K(\cdot)$ is bounded, of bounded variation, has a bounded support and

$$\int K(x)\,\mathrm{d}x = 1.$$

The random function $\hat{f}_n(t)$ may be represented in the form

$$\hat{f}_n(t) = f^{(n)}(t, \theta) + \frac{1}{\sqrt{n}} \int \frac{1}{h_n} K\left(\frac{t-u}{h_n}\right) \mathrm{d}W_n(u), \tag{3}$$

where we denote for a function φ (may be vector function)

$$\varphi^{(n)}(t) = \int \frac{1}{h_n} K\left(\frac{t-u}{h_n}\right) \varphi(u)\,\mathrm{d}u, \tag{4}$$

$$W_n(t) = \sqrt{n}\left(F_n(t) - F(t, \theta)\right).$$

Here F is the distribution function of the random variables X_i and F_n is its empirical version:

$$F_n(t) = \frac{1}{n} \sum_{j=1}^{n} \mathbb{1}_{]-\infty, t]}(X_j).$$

We suppose also that

$$h_n \to 0 \quad \text{and} \quad n h_n \to \infty, \quad \text{when} \quad n \to \infty.$$

It is well-known that in this case

$$\mathbf{E}_\theta \hat{f}_n(t) \to f(t,\theta),$$

i.e. $\hat{f}_n(t)$ is asympotically unbiased for $f(t,\theta)$ at any point of continuity of f, and since

$$n h_n \mathbf{Var}_\theta \hat{f}_n(t) = \mathbf{Var}_\theta \frac{1}{h_n} K\left(\frac{t - X_{n1}}{h_n}\right) =$$

$$\frac{1}{h_n}\int_0^1 K^2\left(\frac{t-x}{h_n}\right) f(x,\theta)\,\mathrm{d}x - \frac{1}{h_n}\left[\int_0^1 K\left(\frac{t-x}{h_n}\right) f(x;\theta)\,\mathrm{d}x\right]^2,$$

we have

$$\lim_{n\to\infty} n h_n \mathbf{Var}_\theta \hat{f}_n(t) = f(t,\theta)\int_0^1 K^2(x)\,\mathrm{d}x = k f(t,\theta),$$

i.e. $\hat{f}_n(t)$ is mean squared consistent for $f(t,\theta)$. More about the properties of $\hat{f}_n$ and its applications see, for example, in Silverman (1986), Bickel & Rosenblatt (1973), Härdle & Marron (1990), Härdle & Mammen (1993).

For a function $\varphi(x,\theta)$ and a random point $\theta_* = \hat{\theta}_n$ (where $\hat{\theta}_n$ is the maximum likelihood estimator for θ) we introduce the notation

$$\varphi_*(x) \;=\; \varphi(x,\theta_*), \quad \nabla\varphi\,(x,\theta) \;=\; \left(\frac{\partial}{\partial\theta_1}\,\varphi(x,\theta),\cdots,\frac{\partial}{\partial\theta_k}\,\varphi(x,\theta)\right)^T,$$

$$\nabla_*\varphi\,(x) \;=\; \left(\frac{\partial}{\partial\theta_1}\,\varphi(x,\theta_*),\cdots,\frac{\partial}{\partial\theta_k}\,\varphi(x,\theta_*)\right)^T.$$

Further for simplicity of notation we will write f instead of $f(\cdot,\theta)$ and f_* instead of $f(\cdot,\hat{\theta}_n)$. For a nonnegative function π we denote by L^2_π the L^2-space generated by the measure with density π. Let $\mathbf{I}(\theta)$ be the information matrix,

$$\mathbf{I}(\theta) \;=\; ||I_{i,j}||_{i,j=1,\cdots,k} \;=\; \int \frac{\partial}{\partial\theta_i} f(x,\theta)\,\frac{\partial}{\partial\theta_j} f(x,\theta)\, f^{-1}(x,\theta)\;\mathrm{d}x.$$

We denote by $\psi_j(x,\theta)$ for $j = 1,\cdots,k$ the coordinates of the vector-function $\psi = (\psi_1,\cdots,\psi_k)^T$,

$$\psi(x,\theta) \;=\; \mathbf{I}^{-1/2}(\theta) f^{-1}(x,\theta)\nabla f(x,\theta).$$

So we have

$$\left(\psi_i(\cdot,\theta),\psi_j(\cdot,\theta)\right)_{f^{-1}} = \delta_{i,j}$$

where

$$(\cdot,\cdot)_{f^{-1}} = (\cdot,\cdot)_{\frac{1}{f}}$$

is the inner product in the space $L^2_{f^{-1}}$. The same relation is true for the coordinates $\psi^*_j(x)\,(j=1,\cdots,k)$ of the vector-function $\psi_* = \psi(x,\hat{\theta}_n)$. For a function $h \in L^2_{f_*^{-1}}$ we put

$$[h]_L\,(x) = \sum_{j=1}^{k} \left(h(\cdot),\psi^*_j(\cdot)\right)_{f_*^{-1}} \psi^*_j(x), \quad [h]^L = h - [h]_L\,. \tag{5}$$

It is clear that $[h]_L$ is the projection of h in the space $L^2_{f_*^{-1}}$ on the linear space $L = L_n$ spanned by a set $\{\psi^*_j(x)\,, j=1,\cdots,k\}$.
As test statistic we suggest a function, which is measuring the normalized deviation of the kernel estimator $\hat{f}_n(x)$ from a modified parametric estimator $f^{(n)}(x,\hat{\theta}_n)$. We will consider these estimators as elements of L^2_π. However we measure the distance between functions from L^2_π with the help of two seminorms $\|\cdot\|_1$ and $\|\cdot\|_2$:

$$\|h\|_1 = \|\,[h]_L\,\|_{f_*^{-1}}, \quad \|h\|_2 = \|\,[h]^L\,\|_\pi.$$

The process

$$\zeta^{(n)}(t) = \sqrt{n}\left[\hat{f}_n(t) - f^{(n)}(t,\hat{\theta}_n)\right] \tag{6}$$

further will be called the observable empirical process. We define

$$\mathbf{T}^{(n)}_1 = \|\left[\zeta^{(n)}(\cdot)\right]_L\|^2_{f_*^{-1}} \quad \text{and} \quad \mathbf{T}^{(n)}_2 = \|\left[\zeta^{(n)}(\cdot)\right]^L\|^2_\pi. \tag{7}$$

For nonnegative functions π and h we put

$$\mu(h) = \mathbf{k}\int h(x)\pi(x)\,\mathrm{d}x, \quad \sigma^2(h) = \mathbf{k}^*\int h^2(x)\pi^2(x)\,\mathrm{d}x, \tag{8}$$

where

$$\mathbf{k} = \int K^2(x)\,\mathrm{d}x, \quad \mathbf{k}^* = \int\left(K*K\right)^2(x)\,\mathrm{d}x.$$

As test statistic we suggest the statistic $\mathbf{T}^{(n)}$

$$\mathbf{T}^{(n)} \;=\; \mathbf{T}_1^{(n)} \;+\; \left\{ \frac{h_n^{-1/2}\left[h_n \mathbf{T}_2^{(n)} - \mu(f_*)\right]}{\sigma(f_*)} \right\}^2 . \tag{9}$$

Now we introduce the assumption under which we plan to investigate the asymptotic behavior of the statistic $\mathbf{T}^{(n)}$.

A1. For the kernel holds $K \in L^1$ and $\int K(x)\,\mathrm{d}x \;=\; 1$.
A2. The least even decreasing majorant $K^*(x)$ of $|K(x)|$ belongs to the L^1-space.
A3. The density function $f(x,\theta)$ as function: $\Theta \to L^2_{f^{-1}}$, is continuously differentiable on the open kernel of Θ. That is there exists such vector function

$$\nabla f = \left(\frac{\partial}{\partial\theta_1} f, \cdots, \frac{\partial}{\partial\theta_k} f\right)$$

that

$$f(\cdot,\theta) \;-\; f(\cdot,\theta_0) \;=\; \langle \nabla f\,(\cdot,\theta_0), \theta - \theta_0\rangle \;+\; \mathbf{r}_{\theta_0}(\cdot,\theta),$$

where $\langle\cdot,\cdot\rangle$ is the inner product in $\mathbf{R}^k$,

$$\|\cdot\|_{f^{-1}} \to 0, \quad \text{when} \quad \theta \to \theta_0.$$

A4. The Fisher information matrix $\mathbf{I}(\theta)$ is positive definite.

2. The kernel smoothing

Assuming A1 and let $K^*(x)$ be the least even decreasing majorant of $|K(x)|$,

$$K^*(x) \;=\; \sup_{|t|\geq x} |K(t)|, \quad x \geq 0.$$

We consider the operators

$$\left[A^h f\right](x) \;=\; \int \frac{1}{h} K\left(\frac{x-t}{h}\right) f(t)\;\mathrm{d}t \;=\; f^h(x), \quad h > 0,$$

$$\left[M f\right](x) \;=\; \sup_{h>0} \int \frac{1}{h} K\left(\frac{x-t}{h}\right) f(t)\;\mathrm{d}t.$$

Suppose that the nonnegative function $\omega(t)$ satisfies the condition

$$\sup_I \frac{1}{|I|} \int_I \omega(x)\ \mathrm{d}x \cdot \frac{1}{|I|} \int_I \frac{1}{\omega(x)}\ \mathrm{d}x \ < \ \infty, \tag{10}$$

where I is an interval, $|I|$ is the length of I.

Theorem 2.1 (see J.B.Garnett (1981)). If $K^* \in L^1$ and weight function ω satisfies the condition (10), then
1. M is a bounded operator on L^2_ω;
2. A^h are uniformly bounded (for $h > 0$) operators on L^2_ω;
3. if $f \in L^2_\omega$, then $f^h \to f$ in the metric of the space L^2_ω, when $h \to 0$.

3. Maximum likelihood estimator

Consider a sample $\mathbf{X} = (X_1, X_2, \cdots, X_n)$ of i.i.d. random variables X_i with density function $f(\cdot) = f(\cdot, \theta), \theta \in \Theta \subset I\!\!R^m$. The maximum likelihood estimator $\hat{\theta}_n$ is a measurable solution of the likelihood equation

$$\nabla_\theta\, L_n(\mathbf{X}, \theta) \ = \ \sum_{j=1}^{n} \nabla_\theta\ \log f\,(X_j, \theta) \ = \ \mathbf{0}. \tag{11}$$

It is well-known (see, for example, Witting & Nölle (1970), Konakov (1978), Greenwood & Nikulin (1996)), that under H_0 and some smoothness conditions on function $f(x, \theta)$ we have

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \ = \ \frac{1}{\sqrt{n}} \sum_{j=1}^{n} I^{-1}(\theta) \nabla_\theta \log f\,(X_j, \theta) \ + \ \mathbf{r}_n, \tag{12}$$

where $\mathbf{r}$ is a random vector such that $\mathbf{r}_n \xrightarrow{\mathbf{P}} \mathbf{0}$ (we shall write $\mathbf{r}_n = o_{\mathbf{P}}(1\!\!1)$). Thus,

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \ = \ I^{-1}(\theta) \int \ \nabla_\theta \log f\,(x, \theta)\ \mathrm{d}W_n(x) \ + \ o_{\mathbf{P}}\,(1\!\!1)\,. \tag{13}$$

It is clear that under smoothness conditions on the functions $f(x, \theta)$

$$\int \ |\, [\nabla_\theta \log f]^h\,(x, \theta) \ - \ \nabla_\theta \log f\,(x, \theta)\,|^2 f(x, \theta)\ \mathrm{d}x \ \to 0,$$

when $h \to 0$. Therefore we deduce from (13) that

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \;=\; I^{-1}(\theta) \int \left[\nabla_\theta \log f\right]_n^K (x,\theta)\ \mathrm{d}W_n(x) \;+\; o_{\mathbf{P}}\left(\mathbb{1}\right). \tag{14}$$

A similar representation holds under the set K_n of nonparametric local alternatives

$$K_n \;=\; \{f:\ f(\cdot) = f(\cdot,\theta_n) + N_n v\left(\frac{\cdot - c}{b_n}\right),\ n = 1,2,\cdots\}, \tag{15}$$

where

$$\theta_n = \theta + n^{-\beta}\psi,$$

ψ is a given vector, $\beta > 0$, v is a given function, N_n, b_n are sequences of positive numbers tending to 0, c is a constant (see e.g. Liero et al. (1998)).

4. The modified empirical process

We consider the smoothed empirical process

$$\xi^{(n)}(t) \;=\; \int \frac{1}{h_n} K\left(\frac{t-x}{h_n}\right)\ \mathrm{d}W_n(x), \tag{16}$$

where

$$W_n(x) = \sqrt{n}\left(F_n(x) - F(x,\theta)\right),$$

and investigate the limiting behavior of the L^2-norm of the projection $\xi_L^{(n)}(t)$ of the process $\xi^{(n)}(t)$ on some finite-dimensional subspace L. For a nonnegative function π we put, using (8),

$$\|\xi\|_\pi^2 \;=\; \int |\xi(x)|^2 \pi(x)\ \mathrm{d}x,$$

$$\mu(f) \;=\; \mathbf{k} \int f(x)\pi(x)\ \mathrm{d}x, \quad \sigma^2(f) \;=\; \mathbf{k}^* \int f^2(x)\pi^2(x)\ \mathrm{d}x.$$

It is well-known that under some appropriate conditions (see Bickel and Rosenblatt (1973) and Hall (1984))

$$\mathbf{P}\left\{\frac{h_n^{-1/2}\left[h_n\|\xi^{(n)}\|_\pi^2 - \mu(f)\right]}{\sigma(f)} \;\le\; x\right\} \;\to\; \Phi(x). \tag{17}$$

Let ϕ be a function such that

$$\phi \in L^2_\pi, \quad \|\phi\|_\pi = 1,$$

and $\xi^{(n)}_\phi$ be the projection of $\xi^{(n)}$ on the one-dimensional subspace generated by ϕ,

$$\xi^{(n)}_\phi(t) \;=\; \int \xi^{(n)}(x)\phi(x)\ \mathrm{d}x \cdot \phi(t).$$

So $\xi^{(n)}_\phi = a\phi$, where the random coefficient a is defined by

$$a \;=\; \int \xi^{(n)}(x)\phi(x)\ \mathrm{d}x \;=\; \int \phi^{(n)}(x)\ \mathrm{d}W_n(x),$$

$$\phi^{(n)}(x) \;=\; \int \frac{1}{h_n} K\left(\frac{x-t}{h_n}\right) \phi(t)\ \mathrm{d}t.$$

Since

$$\mathbf{E}a = 0 \quad \text{and} \quad \mathbf{Var}\,(a) \le \int \left[\phi^{K_n}\right]^2(x)\ \mathrm{d}x,$$

we obtain from Theorem 2.1 the following proposition.

Proposition 4.1. Suppose that weight function π satisfies the condition (10) and for some $C = C(\theta) < \infty$

$$f(x,\theta) \;<\; C\pi(x), \quad \text{for all} \quad x. \tag{18}$$

Then there exists such C_1, which depends only on C, the weight π and kernel K (and does not depend on ϕ), that

$$\mathbf{E}\|\xi^{(n)}_\phi\|^2_\pi \;<\; C_1. \tag{19}$$

It is easily deduced from Proposition 4.1 that

Proposition 4.2. Suppose that the weight function π satisfies the condition (10) and L is a finite dimensional subspace of the space L^2_π. Then under condition (18) such C_2 exists, which depends only on C, weight π, kernel K and $\dim L$, that

$$\mathbf{E}\|\xi^{(n)}_L\|^2_\pi \;<\; C_2, \tag{20}$$

where $\xi^{(n)}_L$ is the projection of $\xi^{(n)}$ on L.

Now we denote by L_n a finite dimensional subspace, $\dim L_n = m$. Let $L_n^\perp$

denote the orthogonal complement of L_n in the space L^2_π and $\xi^{(n)}_{L_n^\perp}$ be the projection of $\xi^{(n)}$ on $L_n^\perp$.

Theorem 4.1. Suppose that the weight function π satisfies the condition (10), L_n is a subspace of the space L^2_π, $\dim L_n$ is fixed, $h_n \to O$, when $n \to \infty$. Then under condition (18)

$$\mathbf{P}\left\{ \frac{h_n^{-1/2}\left[h_n \|\xi^{K_n}_{L_n^\perp}\|^2_\pi - \mu(f)\right]}{\sigma(f)} \le x \right\} \to \Phi(x). \tag{21}$$

Proof. Since

$$\|\xi^{(n)}_{L_n^\perp}\|^2_\pi = \|\xi^{(n)}\|^2_\pi - \|\xi^{(n)}_{L_n}\|^2_\pi,$$

and from Proposition 2 we can conclude that

$$\sup_n \mathbf{P}\left\{ \|\xi^{(n)}_{L_n}\|^2_\pi > y \right\} \to 0, \quad \text{when} \quad y \to \infty,$$

therefore from (17) we obtain (21).

Now we consider a case, when

$$L = L_n = \operatorname{span}\{\psi^*_j(x)\,, j = 1, \cdots, k\}.$$

It must be remind that (see(7))

$$\mathbf{T}^{(n)}_2 = \| \left[\zeta^{(n)}_n(\cdot)\right]^L \|^2_\pi.$$

From the Theorem 4.1. we obtain the next

Theorem 4.2. Suppose that weight function π satisfies the condition (10), the density function $f(x,\theta)$ and kernel $K(x)$ satisfies the conditions **A1** – **A4**. Then under (18)

$$\mathbf{P}\left\{ \frac{h_n^{-1/2}\left[h_n \|\mathbf{T}^{(n)}_2 - \mu(f_*)\right]}{\sigma(f_*)} \le x \right\} \to \Phi(x). \tag{22}$$

5. The observable empirical process

Really we deal with the observable empirical process

$$\zeta^{(n)}(t) \;=\; \sqrt{n}\left[\hat{f}_n(t) \;-\; f^{(n)}(t,\hat{\theta}_n)\right]$$

$$=\; \sqrt{n}\left[\hat{f}_n(t) \;-\; f^{(n)}(t,\theta)\right] \;+\; \sqrt{n}\left[f^{(n)}(t,\theta) \;-\; f^{(n)}(t,\hat{\theta}_n)\right]. \qquad (23)$$

It is clear (see (12), (13), (14)) that under smoothness conditions on the function $f(x,\theta)$

$$\zeta^{(n)}(t) \;=\; \xi^{(n)} \;+\; \eta^{(n)}(t) \;+\; \mathbf{R}_n(t,\theta), \qquad (24)$$

where the process $\mathbf{R}_n(t,\theta)$ weakly converge to zero, when $n \to \infty$, and

$$\eta^{(n)}(t) \;=\; \left[\nabla_\theta f^{(n)}(t,\theta)\right]^T \; I^{-1}(\theta) \int \left[\nabla_\theta \log f\right]^{(n)}(x,\theta) \;\, \mathrm{d}W_n(x). \qquad (25)$$

Let P_{L_n} be the orthoprojector in the space $L^2_{f^{-1}}$ on the subspace

$$L_n \;=\; \mathrm{span}\{\frac{\partial}{\partial\theta_j} f^{(n)}(x,\theta), j=1,\cdots,k\}.$$

We introduce the matrix $R_n(\theta) = ||r_{i,j}||_{i,j=1,\cdots,k}$, where

$$r_{i,j} = \int \frac{\partial}{\partial\theta_i} f^{(n)}(x,\theta)\, \frac{\partial}{\partial\theta_j} f^{(n)}(x,\theta)\, f^{-1}(x,\theta)\;\mathrm{d}x, \quad i,j = 1,\cdots,k.$$

and denote by $R^{-1/2}(\theta)$ its square root. Let

$$\varphi^{(n)}(x,\theta) = \left(\varphi_1^{(n)}(x,\theta),\cdots,\varphi_k^{(n)}(x,\theta)\right) \;=\; R^{-1/2}(\theta)\nabla f^{(n)}(x,\theta).$$

It is obvious that

$$\left(\varphi_i^{(n)}(\cdot,\theta), \varphi_j^{(n)}(\cdot,\theta)\right)_{f^{-1}} \;=\; \delta_{i,j},$$

where $(\cdot,\cdot)_{f^{-1}}$ is the inner product in the space $L^2_{f^{-1}}$. So the system $\{\varphi_i^{(n)}(\cdot,\theta), i = 1,\cdots,k\}$ forms the orthonormal basis (in the metric of the space $L^2_{f^{-1}}$) of the subspace L_n. Therefore

$$\left[P_{L_n} h\right](t) \;=\; \sum_{i=1}^{k} \int h(x)\varphi_i^{(n)}(x,\theta) f^{-1}(x,\theta)\;\mathrm{d}x \cdot \varphi_i^{(n)}(t,\theta), \quad h \in L^2_{f^{-1}}.$$

Now we denote by L_n^* the subspace

$$L_n^* = \text{span}\{\frac{\partial}{\partial\theta_j} f^{(n)}(x, \hat{\theta}_n), j = 1, \cdots, k\},$$

and put

$$\left[P_{L_n^*} h\right](t) = \sum_{i=1}^{k} \int h(x)\varphi_i^{(n)}(x, \hat{\theta}_n) f_*^{-1}\, \mathrm{d}x \cdot \varphi_i^{(n)}(t, \hat{\theta}_n), \quad h \in L^2_{f^{-1}(x,\hat{\theta}_n)}.$$

It seems to be reasonable to expect that operators P_{L_n} and $P_{L_n^*}$ weakly converge (on an appropriate sense) when $n \to \infty$ to the orthoprojector P_L in the space $L^2_{f^{-1}}$ on the subspace L,

$$L = \text{span}\{\frac{\partial}{\partial\theta_j} f(x, \theta), j = 1, \cdots, k\}.$$

For our needs it is sufficiently to prove that the asymptotic distribution of the process $\left[P_{L_n^*}\zeta^{(n)}\right](t)$ is the same as the asymptotic distribution of the process $\left[P_L\zeta^{(n)}\right](t)$. It may be prooved by the usual weak convergence technique. One can verify that the limiting distribution of the statistic $\mathbf{T}_1^{(n)}$ is the χ_k^2-distribution.

REFERENCES

1. Aguirre, N., Nikulin, M. (1994). Chi-squared goodness-of-fit test for the family of logistic distributions, *Kybernetika*, **30**, 214-222.

2. Bickel, P., Rosenblatt, M. (1973). On some global measures of the derivations of density function estimates. *Ann. Statist.*, **1**, 1071- 1095.

3. Garnett, J.B. (1981). Bounded analytic functions. Academic Press, N.Y.

4. Ghost, B., Wei-Min Huang (1991). The power and optimal kernel of the Bickel- Rosenblatt test for goodness-of-fit. *Ann. Statist.*, **19**, 999-1009.

5. Härdle, W., Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.*, **21**, 1926-1947.

6. Härdle, W., Marron, J.S. (1990). Semiparametric comparison of regression curves. *Ann. Statist.*, **18**, 63-89.

7. Hall, P. (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators. *J. Multivar. Anal.*, **14**, 1-16.

8. Konakov, V. (1978). Complete asymptotic expansions for the maximum deviation of the empirical density function. *Theory Probability Appl.*, **28**, 495-509.

9. Liero, H., Läuter, H., Konakov, V. (1998). Nonparametric versus parametric goodness-of-fit. *Statistics*, **31**, 115-149.

10. Greenwood, P.E., Nikulin, M. (1996). A Guide to Chi-squared Testing. *J. Wiley & Sons*, N.Y.

11. Rosenblatt, M. (1975). A quadratic measure of deviation of two - dimensional density estimates and a test of independence. *Ann. Statist.*, **3**, 1- 14.

12. Silverman, B.W. (1986). Density Estimation for Statistics and Data Analysis. *Chapman and Hall*, London.

13. Witting, H., Nölle, G. (1970). Angewandte Mathematische Statistik, *Teubner Stuttgart.*